Algorithms for Tumor Cell Recognition

VENELIN TERZIEV
Black Sea Institute, Bourgas,
BULGARIA

TEODORA PETROVA
Trakia University-Stara Zagora,
BULGARIA

MARIN GEORGIEV
Vitalis Ruse, Ruse,
BULGARIA

Abstract: The structure of a medical image analysis system is considered. A histogram analysis of the brightness distribution of healthy and cancer cells of epithelial tissue is performed. A segmentation algorithm based on the brightness histogram is presented. The model has a set of properties that allow simulating the behavior of real tissue. The requirements for the algorithm for detecting tumor cells on a smear are determined.

Key-Words: Algorithm, Tumor Cells, Basic Research, Medicine, Processing, Technologies, Analysis

Received: April 12, 2024. Revised: December 13, 2024. Accepted: January 16, 2024. Published: May 19, 2025.

1 Introduction

Fundamental research in the field of medicine is an important and significant part of the progress of mankind. One of the prerequisites for this is that modern medical technologies are based on fundamental research in the fields of physics, mathematics, biology, and chemistry because progress in science is unthinkable without combining achievements in different, interdisciplinary scientific fields. Digital image processing is used in all fields of science and technology, including in assessing the morphological characteristics of cells. Smart systems are being intensively developed that are designed for the automatic processing of medical images. The purpose of processing is to improve the image quality for further deciphering of data by the clinician or the automatic recognition system. Automated processing and analysis of medical images is a universal tool for medical diagnostics [1-2].

In addition to distortions, the quality of the image is also affected by the illumination, as well as the contrast of the obtained images. One of the most widely used methods for improving the quality of images is by increasing the contrast. Since these methods redistribute brightness levels within a given dynamic range, their application to color images is associated with some peculiarities [4, 6].

Visual assessment of the morphological characteristics of cells is an integral part of human research. Even with the development of flow cytometry technologies, it is impossible to form a complete picture without analyzing cell morphology. Therefore, the task of automating the process of visual microscopy is urgent. To solve such problems, computer image analyzis is used that uses microscopic images [3].

2 Application of Histogram Analysis to Epithelial Tissue Micrograph

One of the main problems of mathematical modeling of cancer development (as well as of biological systems in general) is the fact that the processes that unfold are of different scales [8, 13]. Interpretation of the image histogram is a very important element when working with microscopic images. Detection and clasification of cell nuclei is a complex task due to the heterogeneous structure and high variability of cells. Another feature of the histological image is the uneven illumination of the studied tissues, which also imposes limitations on the use of traditional approaches.

There are still unknown mechanisms of the processes that occur in the tumor. The model allows newly discovered molecular mechanisms to be easily integrated into it. The model can be generalized to the case when epithelial cells spread across curved surfaces [9, 14]. In fact, this case is the most realistic, since epithelial tissue has a complex topology. The number of epithelial cells that can simultaneously participate in tissue evolution is limited only by the computational power. In order to calibrate the physical parameters that reflect the morphology of the cells, materials from the cytological examination

of patients are studied. Fig. 1 is a micrograph of flat epithelial cells, Fig. 1 a is a micrograph of healthy cells (larger), and Fig. 1 b is a micrograph of cancerous cells (smaller, with a hypertrophied nucleus). For the purposes of the study, the images were converted from color to grayscale images. Their normalized histograms were constructed directly from the images of the cells (Fig. 1 c) which shows the distribution of the brightness of the pixels in the images. Fig. 1 c shows the histograms of the brightness distribution of healthy (in brown) and cancer cells (in blue) in the studied tissue.

Fig. 1b shows that cancer cells are, on average, smaller in size. This can be explained by increased internal pressure and greater tightness in the tumor, which arise from rapid cell division. These results were obtained by direct measurement of cells from micrographs and are qualitatively consistent with recent data on cell morphology measurements using new optical methods [11-12].

Fig. 1a.

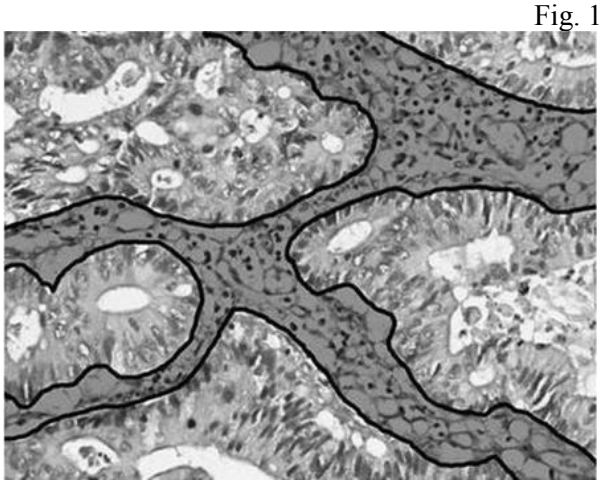

Fig. 1b.

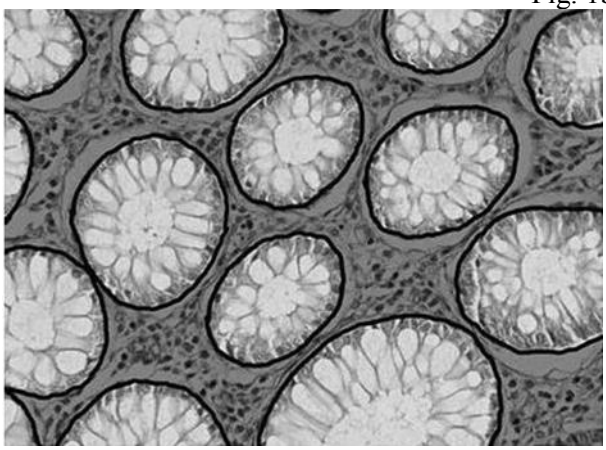

Fig. 1c.

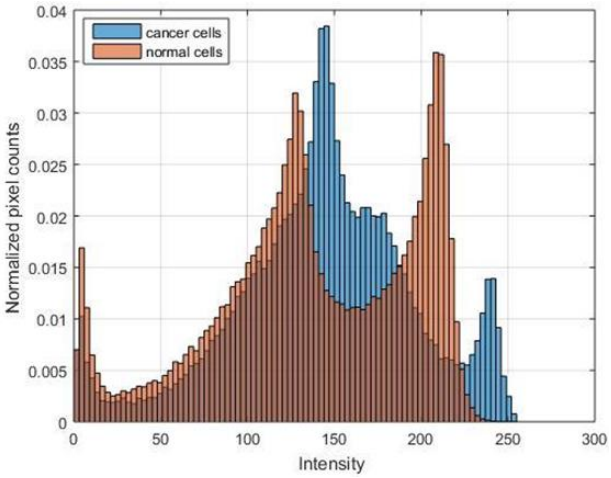

Figure 1. Micrograph showing: a) - healthy epithelial cells, b) - cancer cells, which have a smaller size and an enlarged nucleus; c) - normalized histograms of the distribution of pixel brightness in healthy and cancer cells along the perimeter

Although healthy cells have a more complex shape (Fig. 1a), exotic shapes are not observed when the perimeter varies significantly within a fixed cell area (Fig. 1b). The cells are tightly packed together, forming the continuous two-dimensional surface of the epithelium. The potential energy of the entire epithelial tissue can be considered:

$$E = \frac{1}{2}\sum_{cells}\left(\mu L^2 + \eta\left(A - A_0\right)^2\right), \quad (1)$$

where L and A are the perimeter and area of the cell, respectively. The first term in equation (1) describes the action of forces that tend to reduce the perimeter of each cell, and the second expresses the resistance of the cell to tensile and compressive actions. The corresponding coefficients of elasticity of the medium μ and η are important parameters of the problem. The characteristic value of the area A_0 for healthy and cancerous epithelial cells is determined experimentally. Potential (1) makes it possible to calculate the forces acting on each cell, and hence the evolution of each cell over time, taking into account changes in the local mechanical properties of the medium. The model has a set of properties that allow simulating the behavior of real tissue:

- the possibility of growth of the total number of cells in the system by mitotic division under certain conditions of evolution;
- the ability to move cells in the total mass of the epithelium through the intercalation mechanism;
- calculation of the dynamics of the concentration of a protein that participates in the regulation of the

vital functions of tissues for each cell of the cellular ensemble;

- exchange of chemical signals that occur between neighboring epithelial cells across a common boundary.

The description of the mechanical properties of epithelial tissue must be supplemented by a description of the gene regulation and expression processes that occur in the cell nuclei. For this, a chain of chemical reactions is calculated for each cell. In practice, in a real cell, thousands of reactions occur simultaneously in the nucleus.

3 Operational principles of the algorithm

The proposed algorithm for tumor cell detection consists of two stages, which can be repeated several times for the same image.

In the first stage (based on the study of the brightness histogram of the image), threshold values for the brightness G and the blue fraction fB are selected. In the second stage, the sets of pixels that meet these conditions (primary objects) are examined to see if they can be considered cancer cell nuclei. If the total number of pixels in these fragments is significantly less than the preset number, then the selection of threshold values is considered unsatisfactory and the algorithm is run again.

The first stage is the study of the histogram and selection of threshold values. The first step is to locate the peak associated with the image background, which will later be a reference, both when calculating the optical density and when determining the relative colors for the remaining pixels. Absolute values should be used only if the background peak is not localized. The extreme right peak with a sufficiently small dispersion is chosen as the peak of the background pixels: the standard deviation is less than 10 digits [5, 7]. Then a list of maxima (peaks) is compiled that could correspond to the nuclei of tumor cells. For this, their optical density should be sufficiently large (empirically established limit > 0.6) and the average value of the blue fraction fB = B/(B + G + R) for pixels in this maximum should exceed the similar value for the background by 0.03 (empirically established limit). If several such suspicious maxima are found, they are selected sequentially one after the other (in this case, the second stage of the algorithm is passed), starting from the brightest one. Threshold segmentation is performed in terms of brightness and relative gray content $G < G_{max}$; $fB = B/(B+G+R) < fB_{max}$.

The resulting primary objects are compared with the existing nuclei. If they are not qualified as nuclei, and this is possible in the presence of optically dense leukocytes and large platelets, then the next maximum will be selected. If the tumor peak is not distinguished at all, then threshold values are used, which are no longer based on the current histogram, but on the search history, and if there is none, then on an a priori values. The algorithm scheme is shown in Table 1.

Variable Object Search Algorithm	Obtaining a Brightness Histogram	Localizing a Peak Related to the Background
Setting Absolute Limits	Determining Limits	
Localizing Possible Leukocyte Peaks		
Clasification by Brightness and Relative Blue Share		
End		

Table 1. Schematic of the algorithm for detecting primary objects

The second stage is the examination of the obtained fragments. This part does not depend on the method by which the primary objects were obtained. The algorithm for checking selected objects consists of three cycles. In the first cycle, too large (A>2000 μm^2) and too small (A<11 μm^2) objects are removed. Then the optical density and grayscale characteristics are measured. If there is a search history, then according to the “three sigma” criterion, very bright objects with low optical density are removed and the color verification procedure is called. The remaining objects are conditionally placed in the tumor cell class. In the second cycle, the cytoplasmic reconstruction around the nuclei continues for the remaining objects. The cytoplasm includes a connected set of nearby pixels, which with a high probability (more than 0.95) are not background pixels. The constructed set is rejected if it is too large (more than 2000 μm^2) or the form factor of its outer boundary (the square of the perimeter/area) exceeds a sufficiently large value (more than 50). However, with tight adhesion to normal cells from similar-colored cancer cells, the cytoplasm cannot be separated in such a simple way.

If the brightness histogram was used during segmentation, then, at the end, the consistency of the assumptions and the results obtained are checked. To achieve this, the number of pixels of tumor cells which was computed is compared with the number of pixels in the histogram peak, which is assumed to correspond to them. If the differences are significant (exceed > 50%), and to the left of the peak of the

presumed tumor cells there is one more peak, then the above algorithm is run again [15-19].

4 Conclusion

In conclusion, the model can be generalized to the three-dimensional case and is structurally stable. The number of epithelial cells that can simultaneously participate in tissue evolution is limited only by the computational power. Since newly discovered molecular processes in cells can be easily integrated directly into the model, it can be used to numerically analyze the evolution of real tumors under different initial and boundary conditions.

References:

[1] Batishchev, D.S., Mikhelev, V.M. Infrastruktura vysokoproizvoditel'noy komp'yuternoy sistemy dlya realizatsii oblachnykh servisov khraneniya i analiza dannykh personal'noy meditsiny // *Nauchnye vedomosti Belgorodskogo gosudarstvennogo universiteta. Seriya: Ekonomika. Informatika*. - Belgorod: Izd-vo NIU BelGU, 2016. - S. 88-92.

[2] Belyakov, V.K., Sukhenko, E.P., Zakharov, A.V., Kol'tsov, P.P., Kotovich, N.V., Kravchenko, A.A., Kutsaev, A.S., Osipov, A.S., Kuznetsov, A.B. Ob odnoy metodike klassifikatsii kletok krovi i ee programmnoy realizatsii // Programmnye produkty i sistemy. – 2014. -№ 4 (108). – S. 46-56.

[3] Sokolinskiy, B.Z., Dem'yanov, V.L., Medovyy, V.S., Parpara, A.A., Pyatnickiy, A.M. Automated Sorting Leukocytes in a Blood Smear Using Techniques of Educable Neural Networks and Watershed. *Zdravoohranenie i medicinskaya tehnika* [*Health Service and Medical Technology*], 2005, no. 4 (in Russ.).

[4] Dolgov, V.V., Menshikov, V.V., eds. *Nacionalnoe rukovodstvo po laboratornoy diagnostike [National Guidelines for Laboratory Diagnostics]*, vol. 1, 2.

[5] Medovyy, V.S., Parpara, A.A., Pyatnickiy, A.M., Sokolinskiy, B.Z., Dem'yanov, V.L. Review of Methods of Biomaterial Automated Microscopy. *Klinicheskaya laboratornaya diagnostika [Clinical Laboratory Diagnostics]*, 2006, no. 7, pp. 15–20 (in Russ.).

[6] Medovyy, V.S., Pyatnitskiy, A.M. Robotic Microscopy and information technology to increase accuracy, sensitivity and availability of blood cell analyses. Current microscopy contributions to advances in science and technology *(Microscopy Book Series, Publisher: Formatex Research Center)*, book 5, vol. 1, pp. 775–781, December 2012.

[7] Albertini, M.C., Teodori, L., Piatti, E., Piacentini, M.P., Accorsi, A., Rocchi, M.B.L. Automated analysis of morphometric parameters for accurate defnition of erythrocyte cell shape Cytometry. Part A. 52A:12-18, 2003.

[8] Pratt W.K. Digital Image Processing. *N.Y.–Chichester–Brisbane–Torc. John Wiley & Sons*, 1978.

[9] Panteleev, I., Egorova, O., Klykova, E. Komp'yuternaya mikroskopiya [Computer Microscopy]. Moscow, Tekhnosfera Publ., 2005.

[10] Bikhet, S.F., Darwish, A.M., Tolba, H.A., Shaheen S.I. Segmentation and classifcation of white blood cells. *Acoustics, Speech, and Signal Processing*, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference. On vol. 6. 2000, vol. 4, pp. 2259–2261.

[11] Jiang, K., Liao, Q., Dai, S. A novel white blood cell segmentation scheme using scale-space fltering and watershed clustering. *Proc. of the Second Intern. Conf. On Machine Learning and Cybernetics. Xi'an, 2–5 November 2003*.

[12] Anisimov, B.V., Kurganov, V.D., Zlobin, V.K. Raspoznavanie i tsifrovaya obrabotka izobrazheniy [Image Recognition and Digital Processing]. Moscow, Vyssh. shk. Publ., 1983. 295 p.

[13] Beucher, S. & F. Meyer, 1992. Optical Engineering. *New York: Marcel Dekker Incorporated*, pp: 433-481.

[14] Jambhekar, N. Red blood cells classification using image processing. *Science Research*, 2011, vol. 1, no. 3, pp. 151-154. Date Views 1.04.2018 studyres.com/doc/17754179/red-blood-cells-classification-using-image.

[15] Kokalarov, M. Predetermination of the Uncertainty Budget and Error Analysis of the Photogrammetric Method in Heavy Machinery, *Proceedings of the 30th International DAAAM Symposium "Intelligent Manufacturing & Automation"* (23-26th October 2019, Zadar, Croatia), Volume 30, No.1, pp. 0914-0920, Ed. B. Katalinic / Published by DAAAM International, Vienna, Austria, EU, 2019, ISBN978-3-902734-22-8, ISSN1726-9679, Vienna, Austria. DOI: 10.2507/30th.daaam.proceedings.127.

[16] Kniazieva, V., Artiukh, S., & Kokalarov, M. Measurement of Psychophysiological Indicators in Order to Assess the Professional Qualities of Thermal Power Plants Operators, *2022 XXXII International Scientific Symposium Metrology*

and Metrology Assurance (MMA), (07-11 September 2022), Sozopol, Bulgaria, 2022, pp. 1-4, DOI: 10.1109/MMA55579.2022.9993170

[17] Petrova, T., Petkova-Georgieva, S. Algorithm for controlled computer modeling of the tumor laser therapy process / *Management and Education*. Vol. 16 (6) 2020. pp.51-56.

[18] Terziev, V., Petrova, T., & Georgiev, M. Functional Products for Specialized Nutrition of People Leading an Active Lifestyle. In S. Can & H. Can (Eds.), Evaluating the Effectiveness of Functional Ingredients in Sports Nutrition (pp. 343-372).2025. *IGI Global Scientific Publishing.* https://doi.org/10.4018/979-8-3693-8699-6.ch011.

[19] Marinov, M., Georgiev, M., Petrova, T., Petrov, Zh. Unmanned Aircraft Vechicles as Digital Platforms in Agriculture. Sustainable Perspectives in Business, Economics and Education. Theoretical and Empirical Approaches. *PETER LANG*, 2025, pp. 161-172, ISBN 978-3-631-92917-9.